Available online at www.sciencedirect.com
SCIENCE DIRECT®
Energy and Buildings 38 (2006) 988–996
ENERGY and BUILDINGS
www.elsevier.com/locate/enbuild

# Comparison groups on bills: Automated, personalized energy information

Maithili Iyer [a], Willett Kempton [b,*], Christopher Payne [c]

[a] Lawrence Berkeley National Laboratory, USA
[b] College of Marine Studies and Center for Energy and Environmental Policy, University of Delaware, Newark, DE 19716, USA
[c] Lawrence Berkeley National Laboratory and Center for Energy and Environmental Policy, USA

Received 12 October 2005; received in revised form 9 November 2005; accepted 15 November 2005

**Abstract**

A program called ‘‘Innovative Billing’’ has been developed to provide individualized energy information for a mass audience—the entire residential customer base of an electric or gas utility. Customers receive a graph on the bill that compares that customer’s consumption with other similar customers for the same month. The program aims to stimulate customers to make efficiency improvements. To group as many as several million customers into small ‘‘comparison groups’’, an automated method must be developed drawing solely from the data available to the utility. This paper develops and applies methods to compare the quality of resulting comparison groups.

A data base of 114,000 customers from a utility billing system was used to evaluate Innovative Billing comparison groups, comparing four alternative criteria: house characteristics (floor area, housing type, and heating fuel); street; meter read route; billing cycle. Also, customers were interviewed to see what forms of comparison graphs made most sense and led to fewest errors of interpretation. We find that good quality comparison groups result from using street name, meter book, or multiple house characteristics. Other criteria we tested, such as entire cycle, entire meter book, or single house characteristics such as floor area, resulted in poor quality comparison groups. This analysis provides a basis for choosing comparison groups based on extensive user testing and statistical analysis. The result is a practical set of guidelines that can be used to implement realistic, inexpensive innovative billing for the entire customer base of an electric or gas utility.


*Keywords:* Innovative billing; Energy feedback; Energy efficiency

## 1. Introduction

The utility bill is often the primary basis for customers to understand and analyze their own energy consumption and to draw inferences about their consumption patterns. Research has shown that customers like energy consumption information that provides the means to answer questions such as ‘‘How much money did I save this year?’’ or ‘‘Is my new energy-efficient heater really saving energy?’’ or ‘‘How am I doing compared to people in houses like mine?’’ [1,2]. The University of Delaware (UD), in a cooperative agreement with the United States Environmental Protection Agency (EPA), developed a program to assist utilities that wish to make enhancements to residential customer bills. The enhancement provides information to customers, specifically, comparing each customer’s own energy consumption with that of other homes. The objective of providing such information is to motivate customers to engage in activities leading to energy conservation, whether structural changes to the building or changes in energy management, and to provide a means for evaluating those changes afterwards.

The innovative billing approach is complimentary to, and synergistic with, inspections and energy ratings of buildings, as recently required by the European Commission [3]. Energy inspections and ratings cost in the range of US$ 100–250 and are typically done once per owner (or once per building). Innovative billing is ongoing, costs in the range of 10 cents/year, is not specific as to individual devices or end-uses, gives the occupant a sense of how he or she compares with others in nearby housing, and provides regular feedback to the occupant to evaluate any changes made in energy devices or management.

In this paper, we use residential customer data from an electric utility to evaluate alternative methods for selecting the group of customers with which any one customer is compared. We call this the ‘‘comparison group’’. We use statistical

* Corresponding author. Tel.: +1 302 831 0049; fax: +1 302 831 6838.
*E-mail address:* willett@udel.edu (W. Kempton).

0378-7788/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2005.11.009

techniques to assess the value of different comparison groups selections, and discuss practical use of comparison groups on bills as part of a utility energy efficiency programs. We also address questions of customer interpretation of alternative graphics, and the availability of alternative data sets to utilities.

## 2. Innovative Billing: program description

Innovative Billing is a program that we have developed for use by utilities under US EPA funding, initially using the name ‘‘Energy Star Billing’’[1] (ESB) [4,5]. To implement the program, the utility or its software supplier adds a graph to the bill showing each customer's energy consumption relative to a comparison group. An Innovative Bill is designed so that consumers can understand comparative displays showing their energy consumption and draw valid inferences from them [6]. It differs from comparisons already used on utility bills in that the bill payer is compared with other similar customers for the same month. ‘‘Similar’’ can be defined as similar house size, same street, same sub-division, etc., as discussed in this paper. The improved bill may not be acted upon in majority of the cases. Nevertheless, it is inexpensive and, unlike most energy services, is easily applied to the entire customer base with minimal additional cost. By improving information flow between utilities and their customers, we expect that the program will help to deliver cost-effective, lasting efficiency improvements [7].

Utility bills have the potential for offering energy services which reach a mass market (all customers in the service territory), yet are customized for each customer [8]. Utility bills have historically been constrained by available technology and are molded by the needs and concepts of the utility and its state regulators. Current bill formats result from negotiation among internal utility departments (customer service, data processing, etc.), with some constraints set by public regulators in coordination with consumer advocates and other stakeholders in public regulation [2]. A key motivation for altering existing utility bills stems from research on consumer behavior and energy analysis. This research suggests that a substantial fraction of customers are using their utility bills to analyze consumption, detect changes in consumption patterns, and evaluate the impact of conservation measures [2,9].

### 2.1. *Program approach: comparison groups, stimulating conservation*

Success of customer-focused efficiency programs is to a large extent limited by the design of information provided to the customers. It is easy to fall prey to creating bill information that makes sense to an analyst, but not to the average utility customer. Energy efficiency behavior can be encouraged by clarifying action and consequence through feedback. Several features are important to include in a feedback program in order to strengthen the link between a consumer's action and the consequence of the action. Energy feedback works only in those cases where the consumer is able to recognize the relationship between behavior and outcome. For this purpose, the information should be designed to relate to a comprehensible standard or comparison group [4].

The Innovative Billing program is based on the concept that consumers will use the comparison group information to evaluate their own energy consumption. The University of Delaware conducted a detailed survey on how utility customers interpret and use comparative graphics of their energy use [6]. We expected that customers with high relative consumption would be motivated to investigate conservation measures and observe the impacts of implementing these measures on their Innovative Billing graph over time. In the survey, over 70% of the respondents said they would take energy conservation actions if they received an Innovative Billing graph showing them to be on the high-consumption end (80th percentile) of their comparison group [10,6]. Of course, survey results may not reliably predict what people will actually do. Nevertheless, research has shown that accurate and easy to understand information can motivate consumers to reduce their energy use. Some well-designed pilot energy information programs on billing and continuous metering have achieved savings of up to 13% and costs of conserved energy as low as 1 cent/kWh [1]. However, in a few cases little or no measured savings have resulted from energy information services [11,17,18], a reminder that an iterative design process, pilot implementation, and subsequent evaluation are critical.

The comparison group analysis in this paper is concerned with finding methods of grouping customers that can be postulated to display similar house or energy requirement characteristics. We define each comparison group for customers such that it allows for formation of clusters that are relatively homogeneous within the group, and are suited for meaningful comparisons.

### 2.2. *Importance of good comparison group to customer*

To achieve the most effect from Innovative Billing we want comparison groups that meet two criteria: they are analytically valid comparisons and customers perceive them as valid comparisons. For example, consider a customer in a block of nearly identical houses who is compared with others in their block. If they were much higher or lower in energy use, we expect, as analysts, that it would most likely indicate something about their management of the house or their choice of appliances. In an area with very heterogeneous housing, comparison groups might group together houses of similar construction and appliances, rather than a geographical grouping.

In our study, we are dealing with two principal issues while constructing the comparison groups. The first issue relates to the concept that there are ‘‘natural’’ categories for which comparisons are meaningful.[2] The groupings generally tend

[1] Energy Star is a registered trademark of the US EPA.

[2] Based on several studies on market research and consumer behavior [12,13].

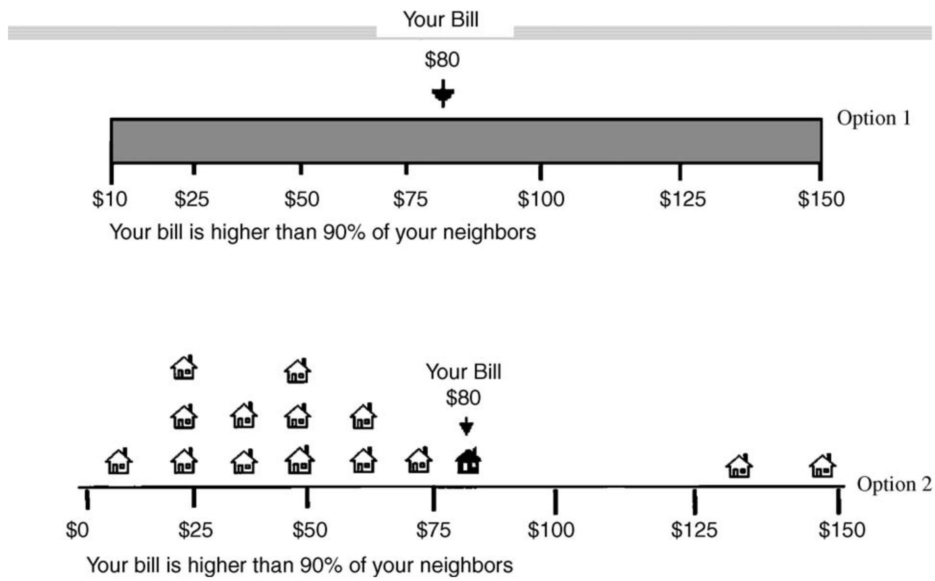


Fig. 1. Two options for graphical display of comparison groups.

to bring together customer clusters that are homogeneous in their characteristics (house characteristics, lifestyle, etc.). Delineating these groups of customers is one aim of statistical grouping. Normally, cluster analysis[3] is used to obtain these ''natural'' groups of people. Since our study deals with a large utility population (a utility customer base being different from a randomly selected set of informants), we use our defined comparison groups as a surrogate for cluster analysis. These methods group customers either geographically or by house type. The other reason for using groupings such as house type of geography is that these groupings are understandable to customers, whereas grouping them by a statistical clustering would be unlikely to be understood or perceived by customers to be valid. Since these groups are not established statistically, it is important to evaluate the groups statistically. We make use of descriptive statistics such as the standard deviation, and measures of skewness and kurtosis for that purpose. The analysis here, we hope, provides sufficiently clear results that they can be used in practice by utilities without the need to statistically re-evaluate alternative comparison group clustering methods.

The second issue we are tackling is that of presenting comparative energy consumption information in ways that minimize misconceptions. As noted earlier, that is addressed in part by studies of readability and comprehension of displays by users. It is also addressed in the statistical analysis of comparison groups by finding criteria for groups that conceptually make sense to users, and that statistically have minimal variation within them.

### *2.3. How can comparison groups be misleading to an energy user?*

Prior research on consumer comprehension of graphical displays of energy comparison groups [10,6] has led us to recommend either of two types of display. As shown in Fig. 1, one option is a bar graph, the second is a distribution graph. We use the term ''bar graph'' to describe the single horizontal bar illustrating the range of values, as on U.S. ''EnergyGuide'' appliance labels. Both examples in Fig. 1 are based on the same data, a house with an US$ 80 electric bill for the current month. These displays have emerged from testing of about a dozen in personal interviews, and subsequently from a set of four tested in a large survey.

When we began testing graphic displays on customers, we expected the bar graph to be the easiest to understand. In fact, our survey revealed that the distribution graph (option 2 in Fig. 1) was correctly understood by the most people (79% versus 63%), and led the largest proportion to say they would take energy efficiency actions (86% versus 77%). Of the four graphs presented on the survey, the distribution graph was also most often reported as the easiest to understand, least often described as ''difficult to understand'', and most often chosen as the graph the respondent would like to receive [6, pp. 50–58]. The bar graph came in second in many of these categories. We recommend the distribution graph,[4] but retain the bar graph as an option in the Innovative Billing program, because some

[3] Cluster analysis is used for estimating groups of similar objects. Similarity is usually based on resemblance coefficients derived from an object's attributes. Applications of cluster analysis could be found in areas where motivations for engaging in any specific activity are being tested [14].

[4] Our recommendation of the distribution graph does rely on design specifics. For example, in one experiment in Norway, focus group participants described the distribution graph as ''childish'', and it was preferred equally to a distribution curve [15]. We attribute this to the relatively large house icons in Wilhite et al.'s graphic (see their Fig. 2), and to the labeling which did not specifically identify the comparison group (e.g. did not say they are houses ''on your block'') and left ambiguous whether each icon represented one house or many.

utilities will not have the computer printing capabilities to produce a distribution graph like that shown in Fig. 1 (option 2).

Comparison groups are a powerful tool for conveying comparative energy use information to consumers. However, care must be taken to ensure that there is no misrepresentation of the consumer's relative position. For instance, a skewed distribution, or the presence of outliers, can pose a potential problem by making an ''average'' consumer look like a ''low-use'' consumer. Fig. 1 illustrates how two types of graphics suggest different conclusions from the same utility data.

The relative position of the customer (''your bill'') in the bar graph in option 1 is in the middle and thus can be perceived by the customer as average. On the other hand, the same information when presented as a distribution graph (option 2) shows that the consumer is a relatively high user of energy. In the underlying data, the skewness is quite high, and that makes the bar graph inappropriate for display. In many of the comparison groups we analyzed, outliers were even more extreme (say, a single house at US$ 250 in the examples of Fig. 1). Even a single extreme outlier exacerbates the problem illustrated in Fig. 1. To reduce this problem, we will also consider criteria for cutting off extreme outliers.

Finally, the bar graph and distribution graph differ in how many customers can be represented on the graph. The bar graph has no upper limit, since individual points are not shown. The distribution graph does have upper limits. If we are to retain the self-explanatory value of each point as a house icon, a graph that might fit in a corner of a utility bill, given a typical distribution, might contain only 20–40 points. If we went to each house as a smaller symbol, say, open squares (with a filled square for the recipient's house), we might be able to increase this to something like 200. Because we consider the house icon an important aid to understanding,[5] we recommend about 30 as a maximum for distribution graphs.

## 3. Logic of alternative comparison groups

What are the logical possibilities for comparison groups? In the following sections we compare alternative methods for creating comparison groups. They divide into basically two methods: house characteristics and proximity. Energy analysts are more familiar with comparisons based on house characteristics, and may tend to first think of comparison groups clustered by size, physical construction, and equipment.

Although house characteristic comparisons are more familiar, there are four reasons to consider geographical groups as an alternative. First, customers know which specific houses they are being compared with and can communicate with others in their comparison group—even if the neighborhood has heterogeneous housing, customers can make inferences such as ''I've got one of the smaller houses here on Maple Street, so I shouldn't be near the top of this graph''. Second, geographical groups will be easier for many utilities to establish, due to reasons of data quality and availability, as discussed later. Third, geographically proximate houses tend to be similar in both housing characteristics and in social characteristics. This third point is discussed in the following paragraph. And fourth, geographical groups are easier to describe to consumers: compare a graph labeled ''Houses on Maple Street'' with a graph labeled ''100–130 $m^2$ residences with a gas furnace and electric water heat''.

Social science and marketing have studied the similarity of geographically proximate households. Proximate individuals tend to report similar behavior and attitudes [16]. In the field of marketing, clustering or grouping of consumers has been historically grounded in the disciplines of sociology and statistical analysis. Market researchers have maintained that consumers could be more effectively grouped in neighborhood-sized markets. Cluster targeting enables marketers to pinpoint locations of people with a customer profile matched to their product. Commonly used marketing units have included census tracts and zip codes [12]. Although no neighborhood is homogenous in all respects, geographical classification works because the differences among the neighborhoods are larger than the differences among households in the neighborhood. One of the key arguments in favor of clustering used by experts in the area of geo-demography is that ''people are all different, but clustering predicts where you can find more of one kind'' [12].

Whether based on proximity or house characteristics, comparison groups should have common meter read intervals. Drawing houses for comparison within the same meter read dates means that comparison houses are always being billed for the same number of days, and that they experience the same weather. This eliminates any need to adjust the data to compensate for differing weather and eliminates the need to use artificial measures such as ''kWh/day'' in order to normalize for billing-days differences. Comparison only within a cycle means that the easiest to understand measure ''dollars this month'' is also analytically valid for comparisons within the group.

In utility data processing jargon, all customers within a ''cycle'' are scheduled to have their meters read on the same day. A typical US utility will read meters and send bills monthly (some utilities read every other month). Given weekends and occasional holidays, they would typically have 20 cycles within months that average 30 days. If we restrict ourselves to comparison groups only within a single cycle, this means that we must draw comparison groups from within subsets of approximately 1/20 of the total residential data base.

If we were choosing geographically clustered comparison groups, this restriction will have little practical effect, because small geographical units are virtually always in the same cycle. On the other hand, if comparison groups are established on the basis of house data, a draw within a single cycle means that we will be restricted to only 1/20 of the total sample within which we can seek houses of similar characteristics.

[5] The house icon was used in the best-understood graphic tested in our survey [6]. However, we have not tested the distribution graph with and without the house icons, so we are not sure how much it contributed to understanding.

## 4. Evaluating the quality of comparison groups: use of statistical indicators

To measure the quality of comparison groups, we needed to develop quantitative measures. Poor quality groups are heterogeneous, that is, they mix very different house types, family demographics, and energy consumption patterns. For utilities, family demographics and housing type may not be available. Hence, we must evaluate comparison groups based on energy use patterns as a proxy for other variables. Thus, we pick comparison groups based on geography or other house characteristics, and evaluate their quality based on the distribution of energy use within the resulting groups.

The goal of these comparisons is to measure the homogeneity of each of these methods of grouping; that is, to see which of these methods of grouping are likely to produce comparison groups which are most similar in their energy characteristics. We take similarity of energy use as an indicator meaning that houses in the comparison group are comparable to each other and thus suitable as a reference group for benchmarking one's own energy use.

Energy consumption of a group of customers is typically a skewed distribution, with a peak below the mean, and a long tail out to the right (to the high consumption values). An example is shown in Fig. 1 (option 2), but this is also true at all scales, from block, to meter book, to cycle, to the entire utility. For customers to get a sense of how they compare with a group, we postulate that a ''good'' comparison group will have the following characteristics: few outliers and low skewness, so one is not comparing mansions with efficiency apartments; smaller standard deviation, indicating that the comparison group consists of similar energy users. As a measure of outliers, we counted a percentage of the customers who were far from the mean. We arbitrarily counted as outliers those points two standard deviations (2S.D.) from the mean and 3S.D. from the mean because they are well-known in statistics; since virtually all the outliers are on the high side, an alternative measure could have been some percentage of points above some multiple of the mean value.

An additional criterion applies if a utility is using a bar graph (Fig. 1, option 1). For bar graphs, a good comparison group should be a flat, rather than a peaked, distribution. This flatness is indicated by higher kurtosis. However, we shall see that there is some conflict between the criterion of high kurtosis and the other criteria we consider desirable.

In summary, we consider a ''good'' comparison group to have: low skewness, few extreme outliers, and a low S.D. For bar graphs, a good comparison group would additionally have high kurtosis.

## 5. Characteristics of our sample customer data

The sample data used for the purposes of analysis were provided by Portland General Electric (PGE) for their service territory. The data comprised about 115,000 accounts, drawn from their approximately 600,000 residential customers. The utility drew out a set of customer data such that they met the criterion of getting a geographically contiguous data set. However, geographical edges of our sample are disconnected from adjacent cycle, street, city, etc. The data base used in the analysis was made anonymous by encoding account number, address, city, and geographical coordinates.

After adjusting the accounts for outliers which included records that did not have a full 24 months of consumption data, or which had unreasonable readings for 1 or more months, the data base was divided into 18 cycles. These data were then used to assess the quality of different comparison groups. Results stated in this paper are based on analysis of data for a single month of moderate weather.

## 6. Quantitative comparison and evaluation of comparison group methods

In this section, we present our analysis comparing criteria for forming comparison groups. Table 1 summarizes the statistics.

### 6.1. Types of comparison groups

As a point of departure, we created one ''comparison group'' that is the entire utility. The measures for this comparison group are shown on the first line of Table 1. By our measures of S.D.,

Table 1
Summary statistics on comparison groups, comparing different methods

| Method | Number of groups | Average *N* per group | Mean (kWh) | Average S.D. | Average skew[a] | Average kurt[b] | Average 2S.D. out/*N* (%) | Average 3S.D. out/*N* (%) |
|---|---|---|---|---|---|---|---|---|
| Utility | 1 | 112296 | 1430 | 904 | 1.6 | 4.6 | 4.7 | 1.5 |
| Cycles | 18 | 6239 | 1436 | 892 | 1.5 | 4.0 | 4.7 | 1.4 |
| Meter book | 298 | 379 | 1482 | 876 | 1.3 | 3.0 | 4.5 | 1.2 |
| Meter book-30 | 3882 | 29 | 1434 | 769 | 0.87 | 1.0 | 4.8 | 1.0 |
| Streets | 3075 | 37 | 1564 | 701 | 0.85 | 0.78 | 3.7 | 0.6 |
| Street-30 | 5639 | 19.9 | 1489 | 690 | 0.73 | 0.50 | 4.0 | 0.6 |
| Construction date | 193 | 377 | 1457 | 852 | 1.35 | 2.94 | 4.5 | 1.3 |
| Floor area | 284 | 368 | 1547 | 907 | 1.29 | 2.84 | 4.5 | 1.2 |
| Fuel | 62 | 755 | 1606 | 672 | 1.21 | 3.45 | 3.1 | 0.96 |
| Housing type, area, and fuel | 777 | 94 | 1667 | 587 | 0.74 | 1.00 | 2.4 | 0.47 |

[a] Skewness (a normally distributed data set will have a skewness of 0).
[b] Kurtosis (normalized kurtosis of a standard Gaussian distribution is 0).

skewness, and outliers, this is the worst quality comparison group examined. As we successively shrink the geographical scope of comparison groups, to cycle, meter book, meter book divided into sequences of 30, and streets, the quality of the comparison groups progressively improves. The best geographical clustering is by street name, dividing streets longer than 30 addresses into groups of 30.

Looking at physical characteristics of the housing yields a somewhat surprising result: when comparison groups are based on single criteria, such as floor area alone, or construction date, the comparison groups are very poor—about the same as the whole data base combined. But when several housing characteristics are combined (housing type, area, and fuel), the resulting comparison groups are high quality, slightly better than streets grouped by 30s.

The following subsections discuss the logic of each type of comparison group, and our quantitative measures of the groups' quality.

### 6.2. Meter book as comparison group

If we are to organize the comparison groups by their location (geographically), there are several ways to do so. We can use the utility's grouping and have comparison groups based on meter book; that is, the entire set of houses covered by one meter reader on one day. In our sample data base, the average meter book includes 379 houses. Since this is too many for a good and easy-to-read distribution graph, we also made a second meter book based grouping by taking sequential groups of 30 within the meter book. This grouping has the advantages of being geographically contiguous and always within the same cycle. It is also very simple for the utility, as they already have their customers organized in this way. However, since few customers know of meter books and cycles, it has the disadvantage that it is not easy to describe to the customer.

Meter book comparison groups offer only a little quality increase over the large heterogeneous groups of cycle or whole utility. Meter book-30 is substantially better in quality. As an example of a way to reduce the problem of outliers in a graph, we consider eliminating outliers from meter book based comparison groups. The following compares two standard deviation and three standard deviation cut-offs, for both a distribution graph and a bar graph. We asked, what would happen if we eliminated points from the graph that were two or three standard deviations from the mean? Would customers be able to make valid inferences from data selected in this manner?

For a distribution graph, taking the mean + 3S.D. as a point where outliers can be cut off provides a reasonable graph with some tail to the right. Mean + 2S.D. seems to cut off high consumption cases. For a bar graph, mean + 3S.D. leaves a very long tail to the right, which means that higher-than-average consumption bar graph viewers would see themselves as low consumers. Mean + 2S.D. comes out better, but in most of the high-outlier meter books we examined, even mean + 2S.D. gives a misleadingly low ''self'' position for the large majority of viewers. We conclude that if meter book is the comparison group, a distribution graph with a cut-off of mean + 3S.D. is recommended. A bar graph is not recommended.

### 6.3. Street as comparison group

Another approach is to use the street address to cluster customers. The simplest way to do this is to put all customers with the same street name for the service address in a single comparison group. Thus, for example, a customer living on Elm Street would get a monthly bill that includes a graph showing their energy use against all other households on Elm Street. If we cluster only by street name, there are about 1925 streets in the data base. This changes substantially when we keep all houses in a comparison group in the same cycle (to control for both weather and number of read days). When we cluster by streets within cycle, we get 3075 street based comparison groups, averaging 37 residences each.

However, some streets have very large numbers of addresses, up to several hundred (especially if apartment blocks are included on the street). These methods have the advantage of very clear meaning to the customer, whether they are being compared with ''all houses on Elm Street'', or with ''100 through 400 Elm Street''.

Statistical analysis of street name groups showed that smaller groups would have to be clustered into larger groups, presumably by combining adjacent streets to build a group of a minimum number of customers. From the available data we saw that 82–95% of customers had a comfortably large comparison group based only on street name. Street and more so, street-30, are higher quality than meter book or meter book-30. Thus, street name appears to be a good basis for geographical comparison groups. It provides reasonable size of groups, the groups are similar, and are easy to describe to the customer. Some streets may need to be combined, or split, to make reasonable sized groups. Other than this modest increase in setup effort, streets appear to be a good basis for geographical comparisons.

### 6.4. Houses with similar characteristics as comparison group

Finally, we can create comparison groups based on the physical characteristics of the house. These characteristics can include attributes such as floor area, house types, and type of heating fuel and air conditioning. This approach of grouping by physical characteristics of the house, of course, requires that the utility has, or is willing to acquire, such data. The groups can then be established on the basis of characteristics that are expected to have the greatest impact on energy consumption. Table 1 shows that comparison groups based on single variables, including construction date, fuel, or floor area, are inferior to the geographical groupings. Next, we used a composite of house characteristics for establishing comparison groups. This produced the best quality of groups, somewhat better than the best geographical comparisons. Specifically, we divided customers into comparison groups as follows:

- By cycle, we had each group within only one billing cycle.
- By housing type, we used the four categories in the data base:
  - single family;
  - mobile home or manufactured home;
  - multi-family.
- By heating fuel, we divided into electric and gas.
- By floor area, we eliminated houses listed as below 23 $m^2$ (250 $ft^2$), and those listed as above 465 $m^2$ (5000 $ft^2$).[6] Then we divided into groups as:
  - 23–92.9 $m^2$ (251–1000 $ft^2$);
  - 93–139 $m^2$ (1001–1500 $ft^2$);
  - then continuing by groups of 46.5 $m^2$ (500 $ft^2$) up to 371.5 $m^2$ (4000 $ft^2$);
  - 372–465 $m^2$ (4001–5000 $ft^2$).

House data, like addresses, are readily understood by customers. However, customers may not immediately see that they are the most important characteristics for analyzing energy consumption. For example, customers may consider it more important to compare with other houses that have two teenage sons, or with houses having people who go south for December. Nevertheless, in both analytical validity and in validity as perceived by customers, comparison groups based on house data have a strong advantage over geographical groups when neighborhoods are highly heterogeneous.

## 7. Practical issues of data availability and quality

This paper has examined the quality of comparison groups organized by meter book, service address, and house characteristics. We found that a combined set of house data provided the highest quality comparison groups, and that street address was also very good. Grouping by meter read sequence in groups of 30 was reasonably good. Unacceptably diverse groupings included the whole utility, a whole cycle, and housing groups divided by a single criterion such as square footage. However, a real utility decision must take into consideration more than just the quality measures for comparison groups. Data availability, cost, and quality control effort may also be considerations in choice of comparison group criteria.

One practical advantage of comparison groups based on service address or meter book is that these data are always available. There is no need to purchase the data and no data are missing (we know of one rural co-op which used pole number to find some customers and did not have service address for a few customers, but this is rare).

We compared several levels of house data. Floor area alone, or year of construction alone, was very poor. So, for reasonable house-type based comparison groups, one would need at least floor area, house type, and heating fuel. Although floor area alone is often available from real estate records, the full data may require purchasing data from a company that compiles it.

Once we move beyond service address, we must also consider missing data. Whatever source is used – existing utility load data, real estate records, or purchased customer dwelling data – some records will be missing, others may contain values that are impossible but not marked as missing. We tabulated both missing (a blank or special ‘‘missing value’’ in the data field) and presumed bad data (e.g. a house listed as 10 $ft^2$ floor area). For example, the floor area of the house was considered ‘‘bad data’’ if it was less than 23 $m^2$ or more than 465 $m^2$ (less than 250 $ft^2$ or more than 5000 $ft^2$). Whether or not data are ‘‘bad’’, that is, obviously incorrect, requires some judgment calls, and making a reliable determination of bad data in a large data base requires a thoughtful analyst and some time becoming familiar with the data. We provide tabulations below (see Table 2) of the missing and bad data as an example, asking the reader to keep in mind that other data bases will have very different characteristics.

Differences across data types are dramatic. Meter book and service address are totally reliable. Consumption, needed for an Innovative Billing comparison, is very good; one could afford to write a message instead of giving a graph on 2% of the customers. However, floor area, with 35% missing, heating fuel, with 58% missing, and date built, with 35% missing, are problematic. From general familiarity with these types of data, we do not think the high proportion of missing is atypical in such utility data bases. For some utilities, this will swing the choice of comparison groups definitively to the address based methods (e.g. service address or meter book-30). Since we find house-type comparison groups to be of high quality, we address methods for dealing with missing data of these types.

In a small utility, missing data may not be a problem. For example, when Traer Municipal Utilities in Iowa decided to implement Energy Star Billing, they used city property records for floor area. Although they had no formal records on heating fuel, their customer service representatives knew the 1200 customer service territory well enough that they personally knew many and made educated guesses for all the remainder. Traer sent a letter to each customer stating the values being used, with a request to correct any incorrect values, if the customer wanted a more valid comparison group. In a larger utility, missing data may be a bigger problem. It may be awkward, or generate complaints, if some customers do not receive an Innovative Billing comparison because their data were missing from some data base.

Table 2
Missing and bad data in the study

| Data type | Missing (%) | Bad data (%) |
|---|---|---|
| kWh | 2 | 0.25 |
| Meter book | 0 | 0 |
| Service address | 0 | N/A |
| Floor area | 35 | 0.5 |
| Date built | 35 | 0 |
| Heating fuel | 58 | 0 |

[6] The original data, collected by a US utility company, were expressed in square feet and the size cut-offs are thus in even square feet. Here we give the SI units first, followed by the original English units. Any information presented to the public should of course be expressed in the appropriate national units.

One solution is to treat missing and bad data as a legitimate value for constructing comparison groups. For example, a set of similar houses with missing data for heating fuel might be divided into comparison groups as follows:

Cycle 1, 100–150 $m^2$, multi-family, missing fuel type.
Cycle 1, 100–150 $m^2$, multi-family, gas.
Cycle 1, 100–150 $m^2$, multi-family, electric.

The comparison groups with one value missing are likely to be more diverse, and thus not as ''high quality'' of comparison groups. However, this method does provide a means of mailing out comparison data for everyone.

## 8. Conclusions

This paper considered two issues: (1) how well do the two most-preferred types of Innovative Billing graphs work for real utility data? (2) Which methods of clustering customers result in the highest quality comparison groups?

The typical distribution of utility customers is a skewed distribution, with some high outliers. As illustrated in Fig. 1, this causes potentially misleading results when comparison groups are displayed on bar graphs. For this reason, the distribution graph is preferred to minimize misleading customers; fortuitously, the distribution graph is also the one that our previous survey showed was both most preferred by customers and most often interpreted correctly. Since some utility printing and data processing capabilities will be limited to the simpler bar graph, we also discussed methods, such as cutting off outliers, to minimize the misleading characteristics of this graph. Bar graphs would have also been more acceptable for comparison groups with high kurtosis; however, from the analysis of billing records for one utility, we found that it was not possible to achieve high kurtosis while simultaneously having high quality on our other indicators.

The question that is central to organizing households into comparison groups is, which methods of choosing comparison groups are of higher quality? And which of the display methods are more suited to the method of grouping? The analysis based on our sample utility data revealed that organizing household groups within the same meter-reading cycle is important and relatively easy—this serves to minimize differences in weather and number of billing days. The highest-quality comparison group results from using a combination of house data; however, these data are not available to many utilities. Reasonable comparison groups result from using street name, or meter book and line-of-march, all of which are readily available data. Other methods of dividing the data base into geographical comparisons, such as by entire cycle or by entire meter book, resulted in lower quality comparison groups. We suspect that neighborhood comparisons additionally encourage communication among utility customers regarding methods of energy conservation, although we did not try to measure this social effect. If so, such communication would confer an additional advantage to the neighborhood comparison groups, as opposed to house data comparison groups, beyond the measures of our statistical analysis.

In short, either house data, street name, or meter book line of march can be used for statistically valid comparison groups. Each type offers some advantages, and any of the three could reasonably be picked for the basis of comparison groups.

## Acknowledgements

This research was funded in part by the United States Environmental Protection Agency under cooperative agreement CX 8244452-01 to the University of Delaware; we are grateful to Sam Rashkin at EPA for conceiving and supporting this program. We are grateful to Portland General Electric for use of their customer data, and to Al Pierce for helping us with its use and interpretation. An earlier version of this material was presented at the ACEEE Summer Study, 23–29 August 1998; participant suggestions by participants there, and suggestions from reviewers for *Energy and Buildings* led to improvements in this article.

## References

[1] W. Kempton, Improving Residential Customer Service through Better Utility Bills, E-Source Strategic Memo, SM-95-1, E-Source, Boulder, CO, 1995.
[2] W. Kempton, L. Layne, The consumer's energy analysis environment, Energy Policy 22 (10) (1994) 657–665.
[3] R. Parnell, O. Popovic Larsen, Developing the home energy report: an everyday householder-centred approach, Energy and Buildings 37 (10) (2005) 1092–1103.
[4] D. Lord, W. Kempton, S. Rashkin, A. Wilson, C. Egan, A. Eide, M. Iyer, C. Payne, Energy star billing: innovative billing options for the residential sector, in: Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, American Council for an Energy-Efficient Economy, Berkeley, CA, 1996.
[5] A. Eide, D. Lord, W. Kempton, Innovative billing options: a tool for improving customer relationships in a restructured utility environment, in: Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, American Council for an Energy-Efficient Economy, Berkeley, CA, 1996.
[6] C. Egan, Comparative energy information and its potential in promoting energy efficiency, Masters Thesis, University of Delaware, xi + 153 pp., Fall 1997.
[7] W. Abrahamse, L. Steg, C. Vlek, T. Rothengatter, A review of intervention studies aimed at household energy conservation, Journal of Environmental Psychology 25 (2005) 273–291.
[8] S. Roberts, W. Baker, Towards Effective Energy Information: Improving Consumer Feedback on Energy Consumption, Report to Ofgem, Contract Reference Copn/spec/2003/16, Center for Sustainable Energy, Bristol, UK, 2003.
[9] W. Kempton, L. Montgomery, Folk quantification of energy, Energy—The International Journal 7 (10) (1982) 817–828.
[10] C. Egan, W. Kempton, A. Eide, D. Lord, C. Payne, How customers interpret and use comparative graphics of their energy use, in: Proceedings of the 1996 ACEEE Summer Study on Energy Efficiency in Buildings, American Council for an Energy-Efficient Economy, Berkeley, CA, 1996 .
[11] M. Harrigan, W. Kempton, V. Ramakrishna, Empowering Customer Energy Choices: A Review of Interaction and Feedback in Energy Efficiency Programs, Alliance to Save Energy, Washington, DC, 1995.
[12] M. Weiss, The Clustering of America, Harper & Row, New York, 1988.
[13] B.G. Englis, M. Solomon, To be and not to be: lifestyle imagery, reference groups, and the clustering of America, Journal of Advertising 24 (Spring (1)) (1995) 13(16).

[14] M.S. Aldenderfer, S.K. Bashfield, Cluster Analysis, Sage Publications, Beverly Hills, CA, 1984.

[15] H. Wilhite, A. Hoivik, J.-G. Olsen, Advances in the use of consumption feedback information in energy billing: the experiences of a Norwegian energy utility, in: ECEEE Summer Study Proceedings, Panel 3, 02, 1999 (available from ECEEE Secretariat, Sveavägen 98, 4 tr, SE-113 50 Stockholm, Sweden, eceee@eceee.org, or online 2005 at http://www.eceee.org/library_links/proceedings/1999/abstract/Panel3/ece99_3_02.-lasso).

[16] J. Beaman, J.J. Vaske, An ipsative clustering model for analyzing attitudinal data, Journal of Leisure Research 27 (Spring (2)) (1995) 168–191.

[17] R. Ling, H. Wilhite, An ethnographic examination of the role of energy efficiency in the sale and purchase of large household appliances in Scandinavia, in: Proceedings, 1992 ACEEE Summer Study on Energy Efficiency in Buildings, vol. 10 (Program Design), American Council for an Energy Efficient Economy, Washington, DC, 1992, pp. 111–120 (Available at: www.aceee.org).

[18] J.K. Dobson, J.D.A. Griffin, Conservation effect of immediate electricity cost feedback on residential consumption behaviour, in: Proceedings, 1992 ACEEE Summer Study on Energy Efficiency in Buildings, vol. 10 (Program Design), American Council for an Energy Efficient Economy, Washington, DC, 1992, pp. 33–36 (Available at: www.aceee.org).